

周报（2014.06.09-2014.06.15）

周一：

用上周搭建的 ontology map 的框架实现将 database、file、web 的 local ontology 映射到 global ontology。具体的实现方法是：在 local ontology 到 global ontology 映射时采用 map 文件进行映射。查询完的 local 结果直接按照 global ontology 的名字将查询结果添加到 global Model 中。后面的工作要详细思考 local 和 global 之间的映射，包括 query 的映射、查询结果的映射、ontology 的映射等，这些都需要好好处理。

周二：

抓取雪花网上的新闻信息。具体的方法如下：新建一个 webwrapper，用于封装这个新的新闻数据源的所有内容。因为目前从该网站上抓取的信息是新闻信息，并且能够抓取到的数据与之前构建的新闻本体基本一致，所以这个数据源采用与之前的新闻本体。雪花网上我们需要抓取的信息是动态网页的信息，与之前抓取的新闻网站的结构不同。通过 firebug 抓取访问网页上新闻的具体的链接。不过这个具体的 url 与查询的公司的股票代码相关，而不是像通常的与查询的 url 与查询的关键字直接相关。例如 http://xueqiu.com/statuses/stock_timeline.json?symbol_id=SH600028&count=15&source=%E8%87%AA%E9%80%89%E8%82%A1%E6%96%B0%E9%97%BB&page=1&access_token=K1dHGFF0uhfKILB8rJwQVR&_t=1402403509747 是查询“中国石化”获得的新闻的 json 数据的 url。这个 url 中 symbol_id=SH600028 中的“SH”是在哪个交易所上市的股票，600028 是股票的代码。正好在之前拿到的股票数据里面可以获得企业的名称和其对应的股票代码。因为雪花网上获得的都是公司的股票信息，所以只能查询到上市公司的信息。整个抓取数据的过程中刚开始没有很清楚地了解要使用的方法，所以走了不少弯路。

周三：

对从 stock file 中获得的信息进行细节的修正。之前没有提取 stock file 中的 enterprise 信息，这个可视化的界面比较散，提取了企业的信息后整个数据看起来更紧密地关联起来。对有的公司有多种股票类型，之前只提取了一个，修正后都可以完整地显示。显示股票代码时，文件中的原股票代码信息是

600028.SH，而界面上显示的 instance 的 url 不能以数字开头，之前没有注意到这个问题，把股票代码信息改为 SH600028，就可以完整地显示。还有其他的一些细节的修正，比如 url 中不能有中文的标点符号和某些英文的标点符号，这些小问题之前也没有注意过，现在都已经修正好。

周四周五：

又好好看了一下 palantir 的介绍视频，感觉我们近期的工作应该主要集中在可视化界面上，目前的版本离 palantir 的可视化界面差得好远。之前可视化界面的部分的代码都是解聪和芯芯写的，这两天看了一下这部分的代码，有些还不是很懂，后面感觉这个项目的前端和后端都应该参与的。这周芯芯和哈哈都在忙着浙一的项目，也没有多少时间跟他们讨论。下周应该会在整个可视化界面上展开工作。

下周工作：

跟芯芯和哈哈讨论，进一步确定在界面的整个布局显示方式。